Model Archive Summary for Suspended-Sediment Concentration at Station 11455315/381436121410401; Cache Slough at South Liberty Island near Rio Vista, CA

This model archive summary details the suspended-sediment concentration (SSC) model developed to compute 15-minute SSC beginning July 16, 2013. This is the second suspended-sediment model developed for the site. The methods used follow U.S. Geological Survey (USGS) guidance as referenced in relevant Office of Surface Water/Office of Water Quality Technical Memorandum 2016.10 (USGS, 2016) and USGS Techniques and Methods, book 3 section C, chapter 4 (Rasmussen and others, 2009). This summary and model archive are in accordance with Attachment A of Office of Water Quality Technical Memorandum 2015.01 (USGS, 2014). Note that sediment samples are associated with two site numbers because the station was assigned a new eight-digit station number after it became a flow station.

Site and Model Information

Site number: 11455315
Site name: Cache Slough at South Liberty Island near Rio Vista, CA (LIB)
Location: Latitude 38°14'34.84", Longitude 121°41'03.43" referenced to North American Datum of 1983, Solano County, CA, Hydrologic Unit 18020163.
Equipment: A YSI EXO2 sonde is installed on a data buoy and turbidity data begins on July 16, 2013.

Model number: 11455315.SSC.WY13.1
Model calibration data period: December 5, 2013 – December 4, 2019
Model application date: July 16, 2013 until superseded
Computed by: Tara Morgan-King, USGS, Sacramento, CA (tamorgan@usgs.gov)
Reviewed by: Anna Conlen, USGS, Sacramento, CA (aconlen@usgs.gov)

Physical Sampling Details and Sediment Data

Discrete, boat-based sample collection for SSC monitoring ideally occurs between 6-12 times per year. Sample collection spans the range of conditions targeting storm events during winter and spring flows as well as summer low flow conditions. Sample collection spanned 7 water years (WYs). A total of 37 sediment samples representative of the cross section were collected during WYs 2014-2020. The width of the cross section is approximately 750 feet (ft) at a breached levee site at the south end of a flood bypass. Sample collection varied year to year, with an average of 5 samples collected each water year. The minimum of 3 samples were collected during WY 2016 and a maximum of 10 samples that were collected during WY 2018.

Sample collection is consistent with approved field methods described in Edwards and Glysson (1999). Sediment samples represent the discharge-weighted concentrations of the cross section. Samples are primarily collected using the equal discharge increment (EDI) method to establish five sampling verticals along the transect. Each of the five sections established using the EDI method represents 20% of the total flow. Samples are obtained at the centroid of each equal discharge section. Due to the tidal nature of the site, the EDI method was used to collect discharge-weighted samples to represent the average cross section because velocities are typically not isokinetic (based on Table 4-5 from TWRI09A4, USGS 2006). When an EDI is not possible, an equal width increment (EWI) method is used to collect samples across the cross section. One EWI was collected on 1/9/2018. A boat-based discharge (Q) measurement was collected using an ADCP immediately before sampling to determine the location of each sampling vertical.

Trained USGS technicians collected samples approximately 30 ft downstream of the station piling at the levee breach using a FISP US D-96 depth-integrated, suspended-sediment bag sampler. The channel cross section is roughly 22 ft deep in the thalweg with a mean depth of approximately 15 ft. The minimum depth is roughly around 6 ft and sampling depths vary with the tides and high water occurring during flood events. Average station velocities ranged from roughly -2.0 to + 3.9 ft/sec. Velocities often exceed 2.0 ft/sec specifically during storm events, however they are commonly less than the minimum isokinetic transit-rate requirement for D-96 samplers (2.0 ft/sec). EDI sampling techniques are preferred in non-isokinetic conditions because they still produce a discharge-weighted sample. Sediment at this station was mostly fines (93% on average from sand/fine analysis) and potential bias of SSC due to non-isokinetic sampling is considered minimal.

Samples were analyzed by the USGS Sediment Laboratory in Santa Cruz, California. All samples were analyzed for sediment concentration (mg/L) by the filtration method and most samples were also analyzed for the percentage of fines (< 0.062 mm). The sand/fine break analysis can be used to identify dataset variability and potential outliers and shows that sediment at this station was composed of mostly fines (93% fines on average). Each vertical from the EDI set was analyzed individually by the lab. Individual analysis of each vertical is important for quality control purposes because of rapidly changing, tidal conditions. It can also help define potential lateral variability within the cross section and identify samples contaminated by bed sediment due to nozzle scooping and/or hitting the bed too hard. The set average SSC of the five verticals represents the cross-sectional average and was used in the calibration model dataset. In rare occasions when the SSC at a vertical was deemed an outlier, a revised average was computed from fewer than 5 verticals. This occurred on 12/12/2014, 8/28/2019 and 10/7/2019 with notes applied to the database.

All sediment data were reviewed and marked as approved in the USGS National Water Information System (NWIS) Water-Quality System database (QWDATA) and made publicly available before being included in the sediment model. Analysis results were stored in the NWIS database. Publicly available field/lab sediment data and metadata can be found at: https://waterdata.usgs.gov/nwis/uv?site_no=11455315.

The regression model is based on 34 concurrent measurements of suspended-sediment concentration water samples paired with time-series turbidity data. Model validation follows the guidelines in Rasmussen and others (2009). Summary statistics and the complete model-calibration dataset are provided in the following sections.

Surrogate Data

Continuous 15-minute turbidity and discharge data were collected and computed by the USGS California Water Science Center and evaluated as possible explanatory variables for SSC. The water quality instrumentation is located on a buoy roughly 150 feet upstream of the cross section at center channel. Turbidity data were measured using a YSI EXO2 sonde roughly 1 meter below the water surface and reported in Formazin Nephelometric Turbidity Units (FNU). Turbidity from the EXO sensor begins on 7/16/2013. Note that at the beginning of the YSI EXO deployment, the data were only reported every 30 minutes, so this is reflective of the computed SSC time-series from July 31, 2013 until July 8, 2014. All surrogate turbidity data were computed, reviewed, and approved before using in the sediment calibration model per USGS guidelines (Wagner and others, 2006). Discharge data were collected, computed, reviewed, and approved by the USGS California Water Science Center and retrieved from NWIS-TS (Rasmussen and others, 2009). Methods to compute discharge follow Levesque and Oberg (2012). The 15-minute discharge time-series data are measured and reported in cubic feet per second (cfs) and are available beginning on 2/1/2015. The 15-minute timeseries data are located at: https://waterdata.usgs.gov/ca/nwis/uv?site_no=11455315.

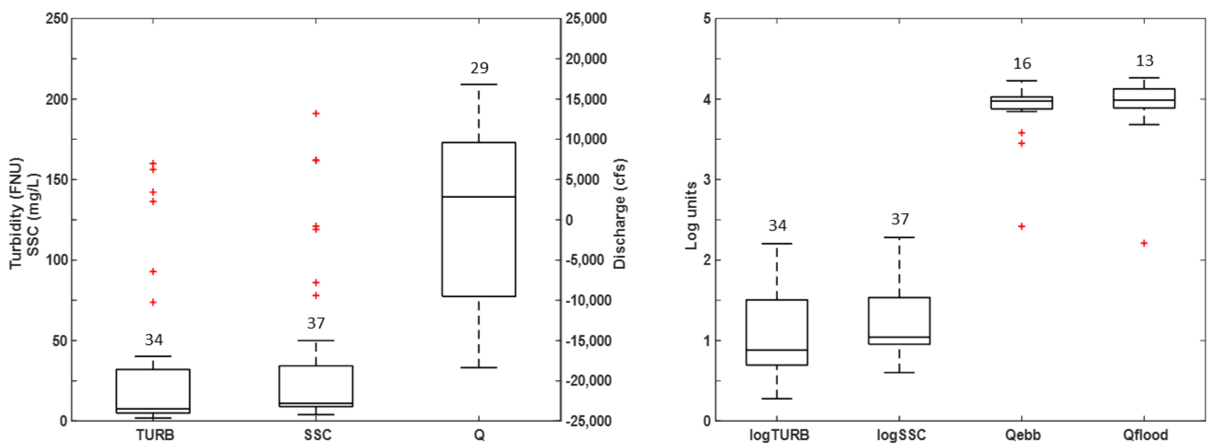
Model Calibration Dataset

The USGS Surrogate Analysis and Index Developer Tool (SAID) was used to pair the surrogate turbidity and discharge data with the discrete sediment data (Domanski and others, 2015). Turbidity and discharge values were paired with each sediment sample observation from a matching max +/- of 15 minutes. The SAID manual is found at https://pubs.er.usgs.gov/publication/ofr20151177.

Of the 37 cross-sectional average sediment samples collected, three did not have corresponding turbidity data due to data gaps in the record, leaving a total of 34 observations in the final calibration dataset. Corresponding turbidity was missing on 12/5/2013, 12/4/2014, and 1/9/2017 so these samples were not used in the dataset. Summary statistics and the complete model-calibration dataset are provided in the following sections.

Regression Model Development

Multiple models were evaluated including simple linear regression (SLR) and multiple linear regression (MLR). The most common estimation technique is SLR, but MLR is an alternate tool for computing SSC when the SLR model standard percentage error (*MSPE*) statistic is larger than 20 percent (Rasmussen and others, 2009). The calibration dataset is composed of 37 SSC, 34 turbidity, and 29 discharge measurements. Boxplots are shown below. Note that due to negative tidal discharge values during the flood tide, ebb and flood values are shown separately with the absolute values shown during flood tides. USGS (2016) *recommends* a minimum of 36 paired observations, however the final dataset was short of that recommendation by two observations due to the missing turbidity values. While the guideline was not achieved, it is equally important that the paired observations span the range of conditions at a site. The dataset achieves this objective as it represents 99% of the conditions - with 40% of the samples collected during the flood tide and 60% collected during the ebb tide.



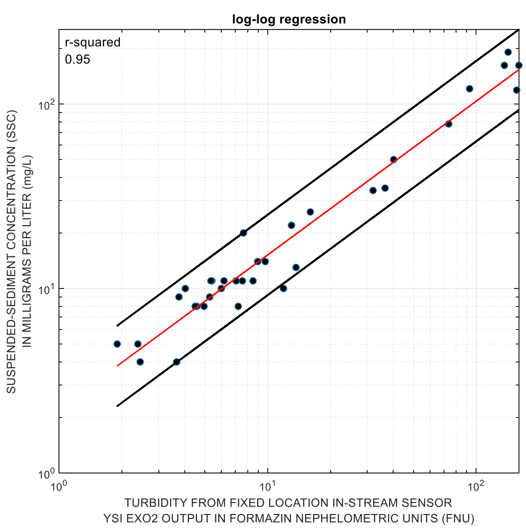
Model diagnostics and plots for model review were output using a combination of Matlab, SAID, and the R environment (R Core Team, 2018). An Rbased application created by the USGS Kansas Water Science Center was also used to produce model statistics and plots for this model archive summary and is available at: [https://patrickeslick.github.io/ModelArchiveSummary/.](https://patrickeslick.github.io/ModelArchiveSummary/) The regression methods used are described in Helsel and Hirsch (2002). Table 3 in Rasmussen and others (2009) shows the best statistical diagnostics to help evaluate the models. The best model was chosen based on residual plots, model standard error, R^2, significance tests (p-values), correlation of explanatory variables, variance inflation factor (VIF), and PRESS (prediction error sum of squares) statistics. Values for the statistics and metrics were computed for various models and are included below along with all relevant sample data and more in-depth statistical information.

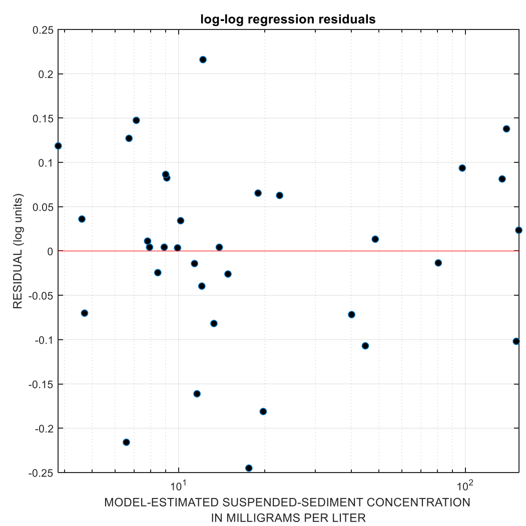
A variety of models were evaluated: Model 1) linear model with one explanatory variable (turbidity), Model 2) $\log_{10}$ transformed model with one explanatory variable (turbidity), Model 3) repeated medians method (Helsel and Hirsh, 2002) using one explanatory variable (turbidity), Model 4) linear model with two explanatory variables (turbidity and discharge), Model 5) $\log_{10}$ transformed model with two explanatory variables (turbidity and absolute discharge), and Model 6) $\log_{10}$ transformed model with two explanatory variables (turbidity and discharge). The number of observations in Model 6 is reduced because the log transformation omits negative discharge values. Diagnostic statistics are summarized below for the six models evaluated. Discharge was not considered further as a second surrogate (in addition to turbidity) because it was not significant as a second variable in the MLR models ($p > 0.05$). The site is variable; sediment can be transported from both upstream and downstream depending on storm tracks and wind patterns. High sediment concentrations coincide with fluvial events, but turbidity is also dependent on wind speeds and transport from nearby flooded shallows and areas that dewater during ebb tides. Thus, discharge (evaluated in Models 4-6) was not considered further in model development.

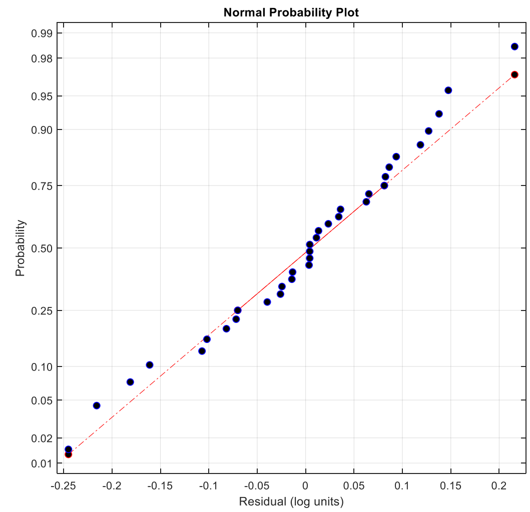
No.	R^2	R^2_a	RMSE	PRESS	MSPE	N	(type)
Model 1	0.94	0.94	12.3	7578	34.2	34	linear
Model 2	0.95	0.95	0.1	0.4	24.9	34	log
Model 3	0.94	0.93	13.3	9241	36.9	34	repeated median
Model 4	0.93	0.93	13.7	10146	37.9	28	multi-linear
Model 5	0.95	0.95	0.11	31.9	24.7	28	ABS multi-log
Model 6	0.94	0.93	0.13	18.9	30.8	15	multi-log

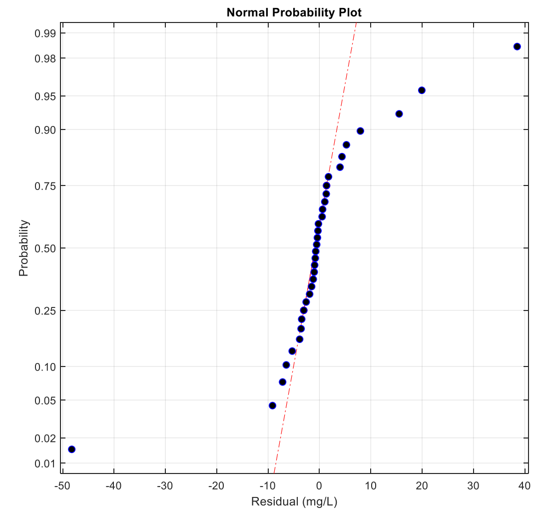
Flagged observations from the SAID outlier test criteria were evaluated. Studentized residuals from the models were inspected for values greater than 3 or less than negative 3. Values outside of the 3 to – 3 range are considered potential extreme outliers. The studentized residuals were reviewed from the SAID output reports and none of the samples were deemed as extreme outliers. The only flags to this dataset was of the Dffits statistic in SAID on 1/22/2016 and 9/13/2018. All 34 observations with concurrent turbidity and SSC were left in the model.

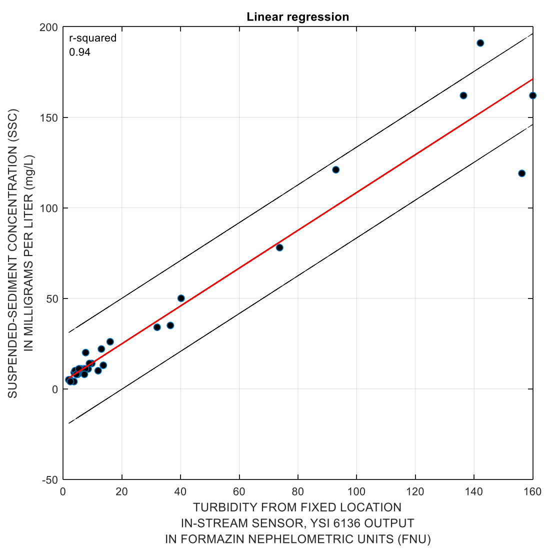
Of the SLR models, the $\log_{10}$ transformed model (Model 2) had the highest R^2, lowest PRESS, and lowest MSPE. Model 2 indicates a more homoscedastic pattern (constant variance) and a more normal distribution compared to the linear model (see the graphs below). Note that while the RMSE and PRESS are not comparable between log and linear models, the predicted SSC was back transformed to get comparable statistics. The back transformed Model 2 RMSE is 12.5 and PRESS is 7563.

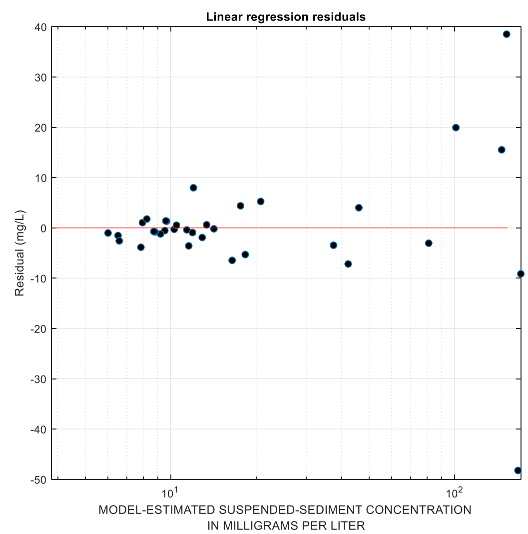





Normal Probability Plot
Probability
Residual (log units)


Normal Probability Plot
Probability
Residual (mg/L)


Linear regression
r-squared
0.94
SUSPENDED-SEDIMENT CONCENTRATION (SSC)
IN MILLIGRAMS PER LITER (mg/L)
TURBIDITY FROM FIXED LOCATION
IN-STREAM SENSOR, YSI 6136 OUTPUT
IN FORMAZIN NEPHELOMETRIC UNITS (FNU)

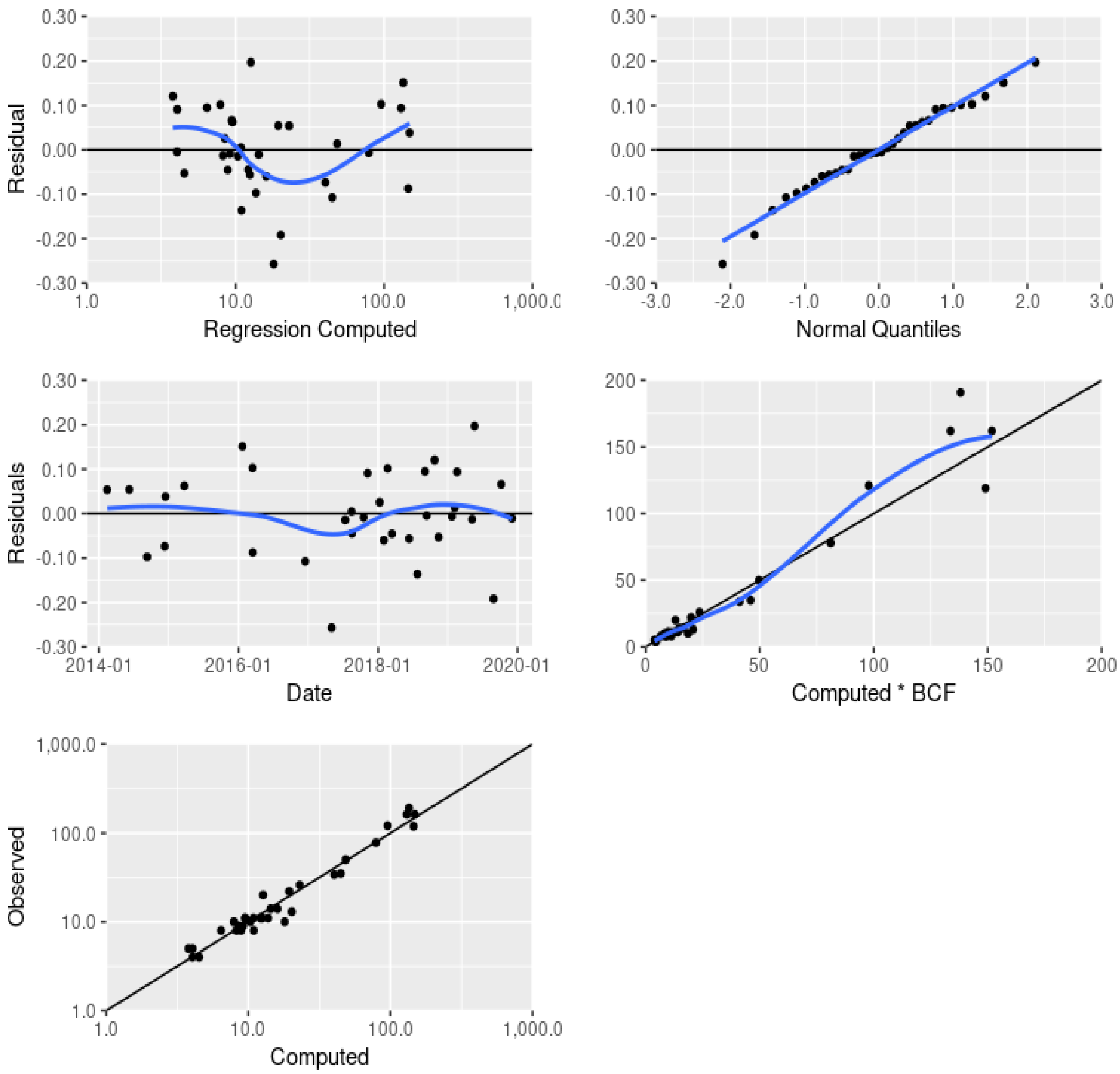

Linear regression residuals
Residual (mg/L)
MODEL-ESTIMATED SUSPENDED-SEDIMENT CONCENTRATION
IN MILLIGRAMS PER LITER

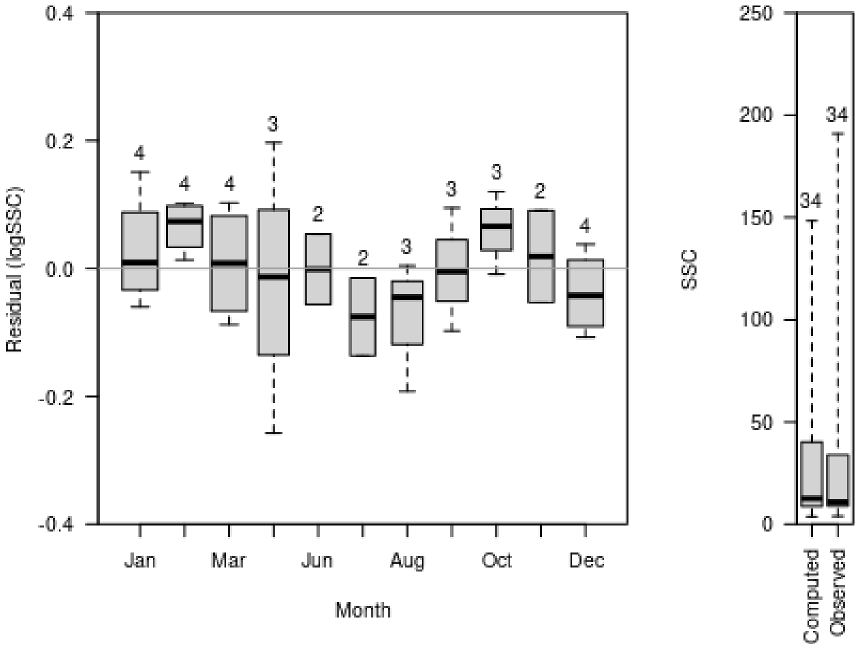
<u>Plots of log_{10}SSC and explanatory variables and residual diagnostic plots</u>

The following Statistical plots were directly generated using a specialized R-Script to follow the specific format guidelines in the USGS technical memo (USGS, 2016). The R-script was developed by Patrick Eslick of the USGS, Kansas Water Science Center (KSWSC) and is located at the following address:

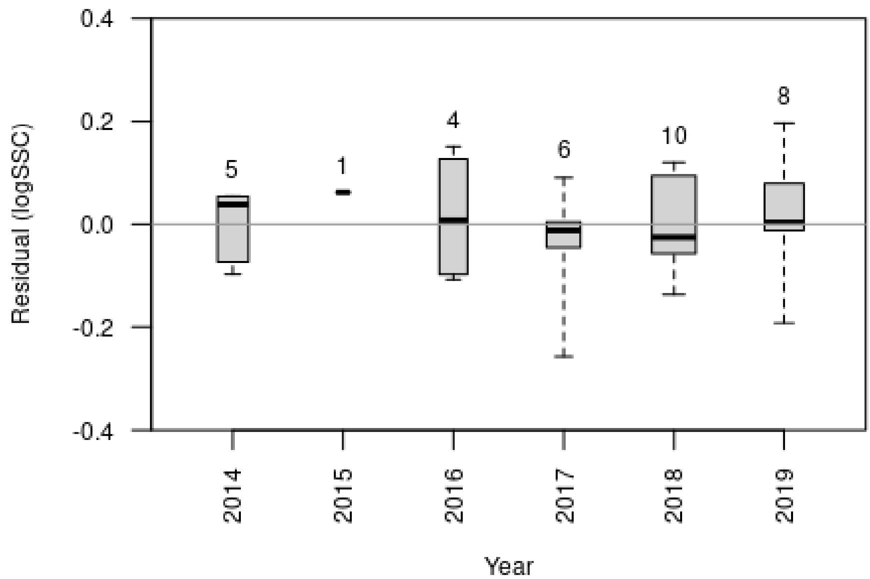
https://patrickeslick.github.io/ModelArchiveSummary/

Statistical Plots



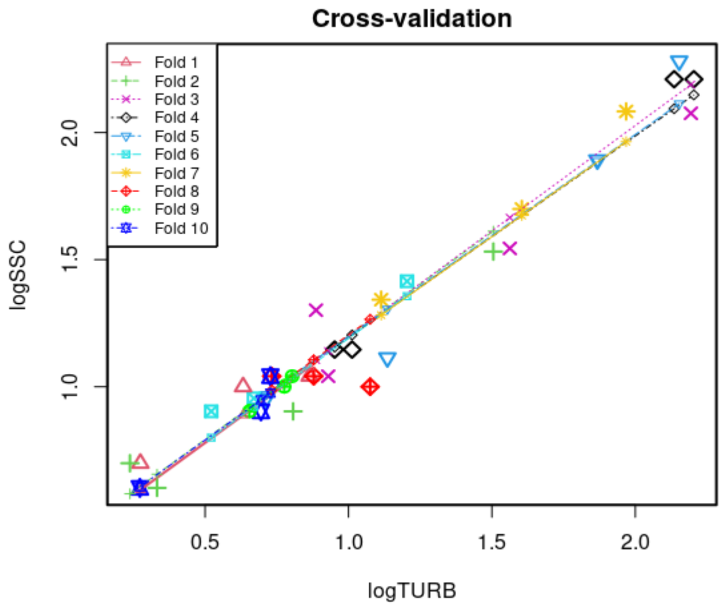


0.4
0.2
0.0
-0.2
-0.4
Residual (logSSC)
4
4
4
3
2
2
3
3
3
2
4
Jan
Mar
Jun
Aug
Oct
Dec
Month
250
200
150
100
50
0
SSC
34
34
Computed
Observed

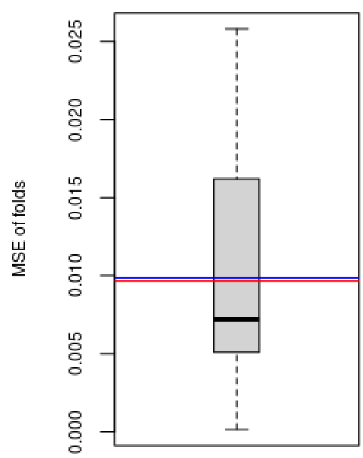


0.4
0.2
0.0
-0.2
-0.4
Residual (logSSC)
5
1
4
6
10
8
2014
2015
2016
2017
2018
2019
Year

The graph below shows a k-fold cross validation with k=10 and the large points represent observations that were left out of each fold and are identified by the color and shape.




```
                 Minimum MSE of folds:  0.000151
                    Mean MSE of folds:  0.009870
                  Median MSE of folds:  0.007210
                 Maximum MSE of folds:  0.025800
 (Mean MSE of folds) / (Model MSE):  1.020000
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

Definitions

```
SSC: Suspended sediment concentration (SSC) in mg/l (80154)
Turb: Turbidity in FNU (63680)
```

MAS App Version 1.0

Model Summary

The final regression model for computing SSC a site 11455315 is a simple $\log_{10}$-transformed model based on 34 concurrent measurements of cross-sectional SSC samples and turbidity collected over approximately 6 water years from February 14, 2014 to December 4, 2019. The log-transformed SLR is shown below with basic model information, regression coefficients, correlation, summary statistics, and Duan's bias correction factor (Duan, 1983):

Linear Regression Model	Coefficient of Determination (R^2)
$log_{10}SSC = 0.348 + 0.834 * log_{10}Turb$	0.95

where

SSC = suspended-sediment concentration, in milligrams per liter (mg/L); and
Turb = turbidity, in formazin nephelometric units

Because SSC was transformed during regression model development, the computed predictions of SSC may be biased and needs to be multiplied by a non-parametric smearing bias correction factor (BCF) which is shown below.

Model	Start date	End date	Linear Regression Model	*BCF*
2	7/16/2013		$SSC = 10^{0.348} \times Turb^{0.834} \times BCF$	1.028

The $\log_{10}$-transformed SLR model can be retransformed and corrected for bias resulting in the following equation:

$$SSC = 2.291\ Turb^{0.834}$$

Parameter	Minimum	Maximum
Turbidity (FNU) entire record	1.0	388
Computed SSC (mg/L)	2	*330/210

*Extrapolation, defined as computation beyond the range of the model calibration dataset, may be used to extrapolate no more than 10 percent outside the range of the sample data used to fit the model. The original maximum computed SSC is 330 mg/L. However, following USGS guidelines, a threshold filter was applied to the time-series limiting the computation above 210 mg/L. Thus, the extrapolated, maximum computed SSC for this model is 210 mg/L.

Suspended-Sediment Concentration Record

The SSC record is computed using this regression model in the USGS National Real-Time Water Quality (NRTWQ) Web site. The complete record can be found at http://nrtwq.usgs.gov/ca.

Model

logSSC = + 0.834 * logTURB + 0.348

Variable Summary Statistics

	logSSC	SSC	logTURB	TURB
Minimum	0.60	4	0.278	1.9
1st Quartile	0.95	9	0.695	4.95
Median	1.04	11	0.881	7.61
Mean	1.25	36	1.08	30.6
3rd Quartile	1.53	34	1.51	32
Maximum	2.28	191	2.2	160

Basic Model Statistics

Number of Observations	34
Standard error (RMSE)	0.107
Average Model standard percentage error (MSPE)	24.9
Coefficient of determination (R^2)	0.952
Adjusted Coefficient of Determination (Adj. R^2)	0.951
Bias Correction Factor (BCF)	1.028

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	0.34828	0.0399	8.72	5.79E-10
logTURB	0.83378	0.0330	25.3	1.00E-22

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.888
E.vars	-0.888	1

Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.176	0.194	0.485

Flagged Observations

Date	Time	logSSC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
1/22/2016	13:21	2.28	2.14	0.138	1.39	1.41	0.135	0.156	0.568
9/13/2018	12:21	0.602	0.818	-0.216	-2.08	2.2	0.054	0.124	-0.527

Model-Calibration Data Set

Observation Number	DateTime	log10SSC	log10Turb	SSC	Turb	Computed logSSC	Estimated SSC	Residual	Normal Quantile	Censored Values
1	2/14/2014 9:40	1.41	1.20	26	16.0	1.35225	23	0.06	0.50	--
2	6/9/2014 9:17	1.34	1.11	22	13.0	1.27707	19	0.07	0.58	--
3	9/11/2014 12:18	1.04	0.93	11	8.5	1.12321	14	-0.08	-0.87	--
4	12/12/2014 10:55	1.53	1.51	34	32.0	1.60325	41	-0.07	-0.76	--
5	12/16/2014 9:51	2.21	2.20	162	160	2.18604	158	0.02	0.26	--
6	3/24/2015 9:31	1.04	0.73	11	5.4	0.958938	9	0.08	0.76	--
7	1/22/2016 13:21	2.28	2.15	191	142	2.14317	143	0.14	1.43	--
8	3/16/2016 11:01	2.08	1.97	121	92.9	1.98918	100	0.09	0.98	--
9	3/17/2016 9:24	2.08	2.19	119	156	2.17749	155	-0.10	-0.98	--
10	12/16/2016 11:05	1.54	1.56	35	36.5	1.65109	46	-0.11	-1.11	--
11	5/3/2017 13:04	1.00	1.08	10	11.9	1.24505	18	-0.25	-2.11	--
12	7/13/2017 11:13	1.00	0.78	10	6.0	0.996436	10	0.00	-0.18	--
13	8/15/2017 9:16	1.04	0.79	11	6.2	1.00717	10	0.03	0.34	--
14	8/17/2017 9:37	1.04	0.85	11	7.1	1.05558	12	-0.01	-0.34	--
15	10/18/2017 13:27	0.95	0.72	9	5.3	0.949937	9	0.00	0.04	--
16	11/8/2017 12:52	0.70	0.38	5	2.4	0.662883	5	0.04	0.41	--
17	1/9/2018 14:18	0.95	0.57	9	3.8	0.827118	7	0.13	1.25	--
18	1/31/2018 12:31	1.15	0.99	14	9.7	1.17211	15	-0.03	-0.50	--
19	2/20/2018 11:11	1.00	0.60	10	4.0	0.852548	7	0.15	1.68	--
20	3/15/2018 11:24	0.90	0.69	8	5.0	0.92751	9	-0.02	-0.41	--
21	6/12/2018 13:31	1.04	0.88	11	7.6	1.08102	12	-0.04	-0.58	--
22	7/26/2018 10:05	0.90	0.86	8	7.2	1.06423	12	-0.16	-1.25	--
23	9/4/2018 12:45	0.90	0.66	8	4.6	0.89883	8	0.00	-0.04	--
24	9/13/2018 12:21	0.60	0.56	4	3.7	0.817915	7	-0.22	-1.68	--
25	10/24/2018 10:36	0.70	0.28	5	1.9	0.580339	4	0.12	1.11	--
26	11/13/2018 11:00	0.60	0.39	4	2.4	0.672187	5	-0.07	-0.67	--
27	1/23/2019 15:18	1.89	1.87	78	73.7	1.90554	83	-0.01	-0.26	--
28	2/7/2019 12:47	1.70	1.60	50	40.2	1.68562	50	0.01	0.18	--
29	2/19/2019 10:54	2.21	2.13	162	136	2.1282	138	0.08	0.67	--
30	5/8/2019 11:41	0.90	0.65	8	4.5	0.891928	8	0.01	0.11	--
31	5/21/2019 10:30	1.30	0.88	20	7.6	1.08505	13	0.22	2.11	--
32	8/28/2019 12:01	1.11	1.14	13	13.7	1.29509	20	-0.18	-1.43	--
33	10/7/2019 12:02	1.04	0.73	11	5.3	0.955012	9	0.09	0.87	--
34	12/4/2019 10:47	1.15	0.95	14	9.0	1.14189	14	0.00	-0.11	--

References

Domanski, M.M., Straub, T.D., and Landers, M.N., 2015, Surrogate Analysis and Index Developer (SAID) tool (version 1.0, September 2015): U.S. Geological Survey Open-File Report 2015–1177, 38 p., https://pubs.usgs.gov/of/2015/1177/ofr20151177.pdf.

Duan, Naihua. 1983. Smearing estimate – A nonparametric retransformation method: Journal of the American Statistical Association. Volume 78-383. 605-610 p.

Edwards TK and Glysson GD, 1999, Field methods for measurement of fluvial sediment: U.S. Geological Survey Techniques of Water-Resources Investigations, Book 3, Chap. C2, 89 p., Available from: https://pubs.usgs.gov/twri/twri3-c2/pdf/TWRI_3-C2.pdf.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources-Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources investigations, book 4, chap. A3, 510 p.

Levesque, V.A. and Oberg, K.A., 2012, Computing discharge using the index velocity method: U.S. Geological Survey Techniques and Methods 3-A23, 148 p, Available at: http://pubs.usgs.gov/tm/3a23/.

R Core Team, 2018, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, Available from: https://www.R-project.org/.

Rasmussen P, Gray JR, Glysson GD, Ziegler AC. 2009, Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data, Book 3 Applications of Hydraulics, Section C, 52 p, Available from: https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf.

Ruhl, C.A., and Simpson, M.R., 2005, Computation of discharge using the index-velocity method in tidally affected areas: U.S. Geological Survey Scientific Investigation Report 2005-5004, 31 p.

[USGS] U.S. Geological Survey, 2006, National field manual for the collection of water quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, Book 9, Chapter A4, Available from: https://pubs.usgs.gov/twri/twri9a4/twri9a4_Chap4_v2.pdf.

[USGS] U.S. Geological Survey, 2014, Policy and guidelines for archival of surface-water, groundwater, and water-quality model applications: Office of Groundwater Technical Memorandum 2015.02, Office of Surface Water Technical Memorandum 2015.01, Office of Water Quality Technical Memorandum 2015.01, Available from: https://water.usgs.gov/admin/memo/SW/sw2015.01.pdf

[USGS] U.S. Geological Survey, 2016, Policy and guidance for approval of surrogate regression models for computation of time series suspended-sediment concentrations and loads: Office of Surface Water Technical Memorandum 2016.07, Available from: https://water.usgs.gov/admin/memo/QW/qw2016.10.pdf.

Wagner RJ, Boulger RW, Jr, Oblinger CJ, Smith BA, 2006, Guidelines and standard procedures for continuous water quality monitors: station operation, record computation, and data reporting: U.S.

Geological Survey Techniques and Methods 1-D3, Available from: https://pubs.usgs.gov/tm/2006/tm1D3/pdf/TM1D3.pdf.